European Economic Review 15 (1981) 137–158. North-Holland Publishing Company

# ENTRY, INNOVATION, EXIT

## Towards a Dynamic Theory of Oligopolistic Industrial Structure*

Partha DASGUPTA

*London School of Economics, London WC2A 2AE, UK*

Joseph STIGLITZ

*Princeton University, Princeton, NJ 08540, USA*

## 1. Introduction

In this paper we explore various factors that influence the relative growth or decline of industries under alternative industrial organizations. Underlying the study is the view that an important ingredient in the performance of an industry over time is the extent to which the opportunities for cost reduction and product improvement made possible by the supply of new inventions are exploited. We focus attention on this by supposing that the flow of new ideas and inventions pertinent to the industry in question is exogenously given; i.e. it is independent of the industry's activities, but that the exploitation of such ideas and inventions involves *innovation*, or *development*, costs. We are then required to analyze the factors that influence the frequency and magnitude of innovations in an industry, and thereby the performance of an industry. But the incentives that firms have for undertaking innovations depend on such ingredients as the growth in demand, the technology of innovations, and the legal structure (e.g. patent rights). In this paper, therefore, we attempt to analyze how these basic ingredients influence the structure and performance of an industry over time. We do this in the context of a model of a single-product industry which can avail itself of a steady stream of inventions and undertake *process innovation*. Innovations, however, involve fixed costs and such costs influence the frequency with which process innovations are undertaken. The analytical device we rely heavily on to conduct the

*This is a revised and expanded version of Section 7 of an invited paper [Dasgupta and Stiglitz (1977)] presented at the World Congress of the International Economic Association on *Economic Growth and Resources* held in Tokyo during August 29–September 3, 1977. We are most grateful to the U.S. National Science Foundation for financial support which made writing this present paper possible, and to Sudipto Bhattacharya, Hans Biswinger, Ashok Guha and Pradeep Mitra for stimulating conversations on the subject.

investigation is the idea of a *steady state*, and for the most part (sections 2 and 3) we study an industry in which there is a kind of 'free-entry' by firms willing to innovate. We postulate that the industry is in a steady state and suppose that firms are profit-maximizing and that they behave in a Cournot manner. We then attempt both to check whether the steady state so postulated is in fact sustained and also to draw out the characteristics of the steady state. For the competitive model with free entry it transpires that the analysis is unusually complicated. We therefore take recourse to certain approximations so as to obtain sharp results. The details of some of the arguments leading to the approximations are provided in the appendix. The body of the paper contains an account of the model and a discussion of the results.

This paper is part of a sequence of theoretical explorations that we have undertaken to study the relationship between the structure of an industry (e.g. the number of firms in the industry, and the degree of concentration) and the nature of innovative activity undertaken within the industry (e.g. the frequency of innovations).[1] The starting presumption in our investigation has been to treat both industrial structure and technical change as endogenous to the model and to view the relationship between them as oligopolistic equilibrium conditions. In our earlier work we analyzed these issues in the context of models postulating a single innovation. Here we study an industry in steady state, which sustains an indefinite sequence of innovations. For analytical tractability we postulate a limited form of free entry, one where any given firm is entitled to innovate and enter the industry only once (e.g. because of the threat of anti-trust suits).[2] Innovation involves a fixed cost. The latest firm to innovate enters with the best-practice technique. The exogenously given flow of new ideas and inventions is modelled by way of a fundamental innovation frontier which represents the trade-off between the percentage rate at which innovation costs increase and the percentage rate at which unit variable cost of production associated with the best-practice technique declines (see fig. 1). The hypothesis that the industry is in a steady state in fact implies that a unique point on this frontier characterizes the rate of technological progress; that is, the rate of reduction in unit variable cost for the industry. However, because innovations involve costs, they are not undertaken continuously, and in fact only a finite number of firms are actually engaged in production at any moment — those whose technologies are sufficiently modern that the marginal costs of production associated with these techniques are less than the market price. Indeed, the innovating firm exits on the date when, due to the postulate of a continual exogenous growth

[1]See Dasgupta and Stiglitz (1977, 1980a, 1980b), and Dasgupta, Gilbert and Stiglitz (1980a, 1980b)

[2]Thus a firm is associated with only one vintage.

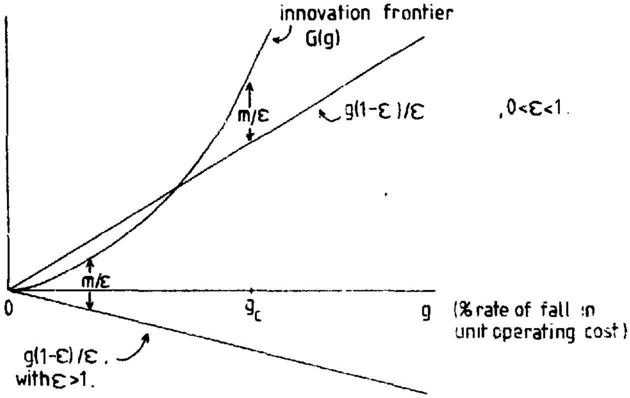



Fig. 1

in demand, the market price falls below the marginal cost of production it encounters. Along the steady state the departing firm is replaced by an innovating firm so that the number of firms engaged in production remains constant. Entry drives the present-value of profits earned by the innovating firm to zero in our approximation, and firms practice mark-up pricing to cover innovation costs. The model, we believe, captures a crucial aspect of the Schumpeterian view of competition, where firms are described as facing a 'perennial gale of creative destruction'. In equilibrium, therefore, the number of active firms is related to the degree of imperfection in competition, the distribution of profits, and in turn to the frequency of innovations, and therefore entry. This explains the title of the paper.

By way of contrast with the oligopolistic industry that results from competition, we analyze the case of an industry controlled by a permanent private monopolist (i.e. protected by entry barriers) in section 5, and also the case of the socially managed industry in section 4. The hypothesis that the industry is in a steady state implies for the model that the rate of technological progress, and therefore the long-run average growth in industry output, is the same for the three industrial organizations under study. The industrial organizations differ, however, by way of the *frequency* with which *innovations* are undertaken and also the magnitude of the innovations when they *are* undertaken. The model highlights the choice an 'industry' has between undertaking minor innovations frequently and major innovations infrequently. The three industrial organizations we study in this paper resolve this choice problem in different manners. Among the conclusions we arrive at are (a) that a monopolist, protected fully by entry barriers, engages in innovations less frequently than is socially optimal, but that when it does, it undertakes more dramatic innovations; and (b) that a competitive industry engages in overly small innovations too frequently if the demand for its product is growing at a fast rate.

## 2. Innovation under competition

In what follows we assume away income effects. We are interested in analyzing steady states. In order to do this we must suppose the (compensated) market demand curve to be iso-elastic. This is admittedly a very special form. But it is the only form which will enable us to study steady states, a central case.

A driving force behind innovation is probably the growth in demand.[3] We accommodate this by postulating a constant rate, $m$, at which the demand curve shifts outwards. To be specific, writing by $Q_t$ total market output at $t$, the market clearing price, $p(Q_t)$, is assumed to be of the form

$$p(Q_t)=\sigma Q_t^{-\varepsilon}e^{mt}, \qquad \sigma, m, \varepsilon>0, \tag{1}$$

where $1/\varepsilon$ is the elasticity of demand.

It should be of no surprise that a dynamic market equilibrium is difficult to characterize in general. For this reason we focus attention on steady state equilibria. Along a steady state innovations occur at fixed intervals of length, $T$, and at all times there is a fixed number of active firms in the industry. In what follows we shall suppose there to be a perfect capital market with the rate of interest $r$ $(>0)$.

Consider an instant in time when a firm innovates and enters. Without loss of generality, let this instant be $t=0$. The innovating firm incurs innovation cost $F_0$ and enters with the best-practice technique at which the marginal cost of production is $C_0$. There are $n$ firms in existence and the previous innovation, by hypothesis, occurred at $t=-T$ when the innovating firm entered with production technique $C_{-T}$ and incurred innovation cost $F_{-T}$. Suppose g is the rate of technical progress (i.e. the percentage rate of fall in unit operating cost). It is, like $n$ and $T$, endogenous to the model. If at $t=0$ the entering firm innovates with the technique $C_0$, then along the steady state, $C_{-T}=C_0e^{gT}$. Let $h$ denote the percentage rate of increase in innovation costs. It too is endogenous to the model and it is linked to $g$ via the innovation frontier which we shall define presently. But since $F_0$ is the cost of innovation at $t=0$, the previous innovator, along the steady state, incurred $F_{-T}=F_0e^{-hT}$ as its innovation cost.

There are $n$ active firms. At $t=0$ — the date of the latest innovation — the $i$th youngest firm (with $i=0,1,\dots,n-1$) in existence operates with technique

[3] The late Jacob Schmookler, in a series of writings, laid great emphasis on this. E.g. see Schmookler (1962).

$C_0e^{igT}$, having incurred the innovation cost $F_0e^{-ihT}$, at the date of entry $t=-iT$. Along the steady state under study the next innovation occurs at $t=T$, the innovating firm entering with technique $C_0e^{-gT}$ and incurring innovation costs amounting to $F_0e^{hT}$. At $t=T$ the oldest firm $(i=n-1)$ will find it most profitable to exit, keeping the total number of firms in operation at $n$. Then, at any instant of time one will observe $n$ firms in operation, each using a different technique of production, where these techniques can be ranked unambiguously in terms of their vintage. The firms, acting non-cooperatively, will of course, not be sharing the market equally, since their operating costs are different. Furthermore, each firm is active for precisely $nT$ years. Thus, in particular, the firm innovating at $t=0$ will exit at date $nT$. For analytical simplicity we shall suppose that a firm can enter the industry at most once. Thus, the firm innovating at $t=0$ will have reached its decision based on an expected survival period of $nT$.

This, in outline, is the characteristic of the industry we shall study. We now proceed to discuss the analytical details which ensure that such a steady state can arise as an intertemporal non-cooperative equilibrium. In what follows we suppose that firms entertain Cournot conjectures regarding others and choose their dates of entry, and the flow of their output during their lifetime, so as to maximize the present value of their net profits. It is in an innovating firm's choice of the timing (and therefore, the magnitude) of the innovation that we are attempting to capture the R&D problem and the forces that determine whether what is essentially a competitive industry is growing or declining relative to the rest of the economy. Thus, at any date $t$, let $C_t$ denote the best-practice technique available *in principle*; that is, the technique with the lowest operating cost known at $t$. The cost of innovation at $t$ is $F_t$. This is the fixed cost required to develop the existing best-practice technique.

We now proceed to describe the innovation frontier. For simplicity of exposition assume for the moment that time is measured in discrete intervals, with the length of an interval being denoted by $\theta$. One can now imagine that from the point of view of the industry in question there is an exogenously given flow of ideas and basic inventions which ensure that, for a given pair $(C_t, F_t)$, there is a set, $\phi(C_t, F_t)$, of feasible pairs of operating and innovation costs from which choice at $t+\theta$ can be made.

Suppose that the efficient points in the set $\phi(C_t, F_t)$ are given by the locus

$$(F_{t+\theta}-F_t)/F_t=\theta G((C_t-C_{t-\theta})/\theta C_{t+\theta}), \tag{2}$$

where $G$ is an increasing and strictly convex function, independent of time, and with the limiting properties $G(0)=0$ and $G(\infty)=\infty$. If we now move to continuous time by letting $\theta\to 0$, the innovation frontier characterized by (2) reduces to the form

$$(\mathrm{d}F_t/\mathrm{d}t)/F_t = G(g_t), \tag{3}$$

where

$$g_t \equiv -(\mathrm{d}C_t/\mathrm{d}t)/C_t. \tag{4}$$

As we are postulating a steady state, $g_t$ is independent of time and so the innovation frontier reduces to the form

$$(\mathrm{d}F_t/\mathrm{d}t)/F_t \equiv h = G(g). \tag{5}$$

The innovation frontier $h = G(g)$ is drawn in fig. 1. We shall note below that if a steady state exists, the parameters characterizing the industry determine uniquely the point on the innovation frontier which sustains the steady state. That is to say, if a steady state exists it is sustained by a unique rate of technical progress, g, which is treated as a parameter by the firms in the industry. What is chosen by firms is the frequency, and therefore the magnitude of innovations; that is, the frequency and magnitude with which this flow of inventions characterizing the technical progress are exploited.

We turn now to the output decision of an active firm. Assume that there are $n$ active firms at any date. Consider the $i$th youngest firm along a steady state. There is no loss of generality if we merely compute its short-run profit maximizing output policy. Let $Q^i$ denote its output at an arbitrary date and let $\hat{Q}_i$ denote the output of the remaining firms in the industry (i.e. $\hat{Q}_i = \sum_{j \neq i} Q^j$). If $C^i$ is the unit operating cost it faces, it must choose $Q^i$ with a view to maximizing $[p(Q^i + \hat{Q}_i) - C^i]Q^i$, which, on using (1) yields as the first-order condition

$$1 - \varepsilon Q^i/Q = C^i/p(Q), \tag{6}$$

where

$$Q = \sum_{i=0}^{n-1} Q^i. [4] \tag{7}$$

Suppose as before that at $t = 0$ an innovation occurs, and let $C_0$ be the unit operating cost the innovating firm faces.[5] Then we may sum (6) to

[4]We need merely to study the first-order condition. The reader will be able to confirm subsequently that if $F_0$ is small and if $m$ is not too small an equilibrium exists, even when demand is inelastic ($\varepsilon > 1$). In particular, we shall note that if $F_0$ is small, then equilibrium $n$ will be large, and in fact that $n > \varepsilon$. Inelasticity of demand can readily be accommodated into the theory of oligopoly with free entry. For details, see Dasgupta and Stiglitz (1980a).

[5]The choice of $t = 0$ as an innovating date is of course without significance. Nothing hinges on it.

obtain the condition

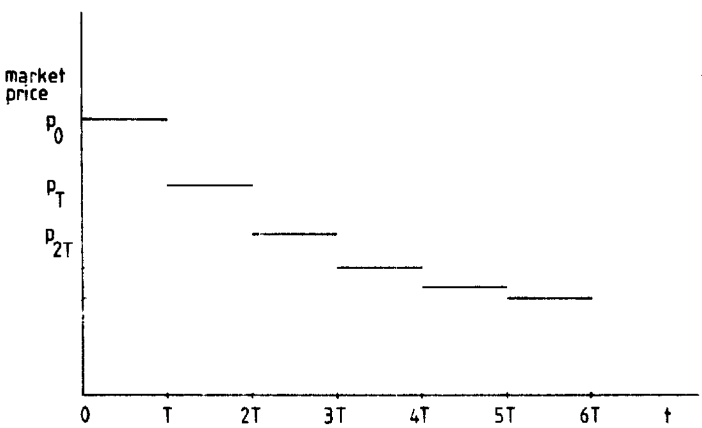
$$n-\varepsilon=C_0\sum_{i=0}^{n-1}e^{igT}/p. \tag{8}$$

From (8) it follows that the price of the product remains constant during the interval between two adjacent innovations [i.e. for $t\in(0,T)$]. But with market price constant, demand, and therefore sales, grows at the rate $m/\varepsilon$ [eq. (1)]. Now at $t=T$, eq. (8) takes the form

$$n-\varepsilon=C_0e^{-gT}\sum_{i=0}^{n-1}e^{igT}/p. \tag{9}$$

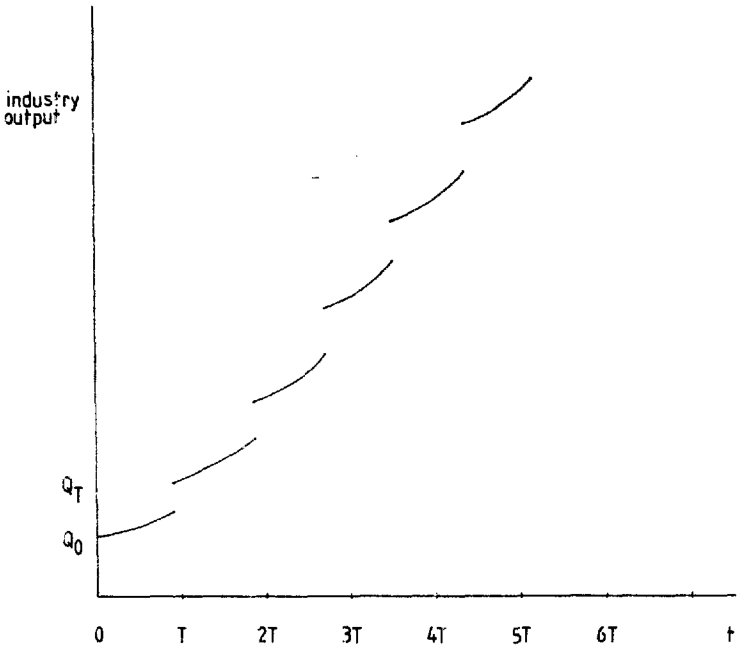
It follows from (9) that at the next innovation date, i.e. $t=T$, market price falls discontinuously by the factor $e^{-gT}$. But this means, by (1), that industry sales must increase discontinuously at $t=T$ by the factor $e^{gT/\varepsilon}$. Since we are considering a steady state, these features are repeated at each innovation date $iT$ $(i=0,1,2,\ldots)$; see figs. 2 and 3.

In order to characterize the oligopoly equilibrium in some detail we need to make certain types of approximations. In what follows we shall suppose that the fixed cost of innovation, $F_0$, is small relative to industry sales at $t=0$. It is simplest to see what this assumption implies if we consider the




$(p_{iT}=p_0e^{-igT}$ and $p_t=p_{iT}$ for $iT\leq t<(i+1)T$; $i=0,1,2,\ldots)$

Fig. 2




$(Q_{iT} = Q_0 e^{i(m+g)T/\varepsilon}$ and $Q_t = Q_{iT} e^{(t-iT)m/\varepsilon}$ for $iT \leq t < (i+1)T$ ; $i = 0,1,2,\ldots)$

Fig. 3

limiting case $F_0 = 0$. In this case the market is perfectly competitive, with $p_t = C_t$ at all $t$ and innovations occur continuously: that is, $T = 0$. In this limiting case the number of firms, $n$, will be infinite.

It follows that if $F_0$ is 'small', then along a steady state $n$ is 'large' and $T$ is 'small'. A firm in equilibrium will operate so long as its cost of production does not exceed the market price. Since we are supposing that $F_0$ is 'small' we are entitled to approximate and suppose that the innovating firm at $t = 0$ will exit at the date market price falls to its marginal cost of production, $C_0$. But this date along the steady state is $nT$. Thus, if $p_0$ is equilibrium product price at $t = 0$, we have also the condition

$$p_0 e^{-gnT} = C_0. \qquad (10)$$

We shall note below that the steady state value of $g$ is a finite positive number. Write $v \equiv nT$, a representative firm's lifetime. Then (10) says that as $F_0 \to 0$, $v \to 0$ (even though $n \to \infty$).

Along the market equilibrium the frequency of innovations must be sufficiently fast that it does not pay firms to enter. Given our approximation we may suppose that the present-value of profits earned by an innovating firm is zero. We now proceed to construct the 'zero-profit' condition for the firm innovating at $t=0$.

Now, at any date, the $i$th youngest firm's flow of profit can be expressed as

$$p(Q)Q^i - C^i Q^i = p(Q)Q^i - p(Q)[1-\varepsilon Q^i/Q]Q^i = p(Q)\varepsilon Q^{i^2}/Q. \tag{11}$$

Therefore, if at $t=0$, the date of the innovation in question, aggregate output is $Q_0$ and market price is $p_0$, then on using (6) in (11) we obtain that the profit flow at $t=0$ to the latest innovator is

$$p_0(1-C_0/p_0)^2 Q_0/\varepsilon. \tag{12}$$

During the interval $(0, T)$ the market price remains constant (see fig. 2), but output grows at the rate $m/\varepsilon$ (fig. 3). Therefore, the present discounted value of the flow of profits to the innovator during $(0, T)$ is, by (12),

$$p_0(1-C_0/p_0)^2 \frac{Q_0}{\varepsilon} \int_0^T e^{-(r-m/\varepsilon)t} dt = \frac{p_0(1-C_0/p_0)^2}{\varepsilon(r-m/\varepsilon)} Q_0(1-e^{-(r-m/\varepsilon)T}). \tag{13}$$

At $t=T$ market price falls discontinuously by the factor $e^{-gT}$, leading to a discontinuous increase in aggregate sales by the factor $e^{gT/\varepsilon}$. Therefore, aggregate sales at $T$ is

$$Q_T = Q_0 e^{(m+g)T/\varepsilon},$$

and market price at $T$ is

$$p_T = p_0 e^{-gT}.$$

Using this in (12) implies that the instantaneous flow of profits at $t=T$ to the firm that innovates at $t=0$ is

$$\frac{p_0 e^{-gT}}{\varepsilon}(1-C_0/p_0 e^{-gT})^2 Q_0 e^{(m+g)T/\varepsilon}.$$

During the period $(T, 2T)$ price remains constant and aggregate sales increase at the percentage rate $m/\varepsilon$. Therefore, as in (13), we may express, from the vantage point of $t=0$, the present discounted value of the flow of profits during $(T, 2T)$ to the firm innovating at $t=0$ as

$$\frac{p_0}{\varepsilon}\mathrm{e}^{-(g+r)T}(1-C_0/p_0\mathrm{e}^{-gT})^2 Q_0\mathrm{e}^{(m+g)T/\varepsilon}[1-\mathrm{e}^{-(r-m/\varepsilon)T}]/(r-m/\varepsilon),$$

and, in general, the present discounted value of the flow of profits during the $i$th period [i.e. during $((i-1)T, iT)$] for the firm innovating at $t=0$ is

$$\frac{p_0}{\varepsilon}\mathrm{e}^{-(g+r)T}(1-C_0/p_0\mathrm{e}^{-igT})^2 Q_0\mathrm{e}^{i(m+g)T/\varepsilon}[1-\mathrm{e}^{-(r-m/\varepsilon)T}]/(r-m/\varepsilon).$$

There are, in all, $n$ periods to consider. Summing these $n$ present-values we obtain the present discounted value of the flow of profits to the firm innovating at $t=0$, which on equating to $F_0$, yields the zero-profit condition. Thus we have

$$\frac{[1-\mathrm{e}^{-(r-m/\varepsilon)T}]}{\varepsilon(r-m/\varepsilon)}\sum_{i=0}^{n-1}\mathrm{e}^{-i(g+r)T}\mathrm{e}^{i(m+g)T/\varepsilon}\{1-C_0/p_0\mathrm{e}^{-igT}\}^2=F_0/p_0Q_0. \tag{14}$$

In fact one can simplify (14) on the supposition that $F_0$ is 'small' so that $T$ and $v$ are 'small'. In the appendix we shall confirm that for small $T$ and $v$, eq. (14) reduces to

$$\frac{g^2v^4}{2\varepsilon}\left[\frac{(m+g)}{\varepsilon}-r\right]\simeq F_0/p_0Q_0. \tag{15}$$

It follows that in order to ensure the existence of a steady state market equilibrium we must suppose that $(m+g)/\varepsilon>r$ — i.e. that innovation is sufficiently profitable. We suppose this to be the case.[6] Furthermore, since the date of innovation, $t=0$, under study has been arbitrarily chosen, eq. (14) says that the ratio of innovation cost to market revenue is the same at all innovation dates.

If we now appeal to (1) and (10), eq. (15) reduces to

$$\frac{g^2v^4}{2\varepsilon}\left[\frac{(m+g)}{\varepsilon}-r\right]\simeq\frac{F_0C_0^{(1-\varepsilon)/\varepsilon}}{\sigma^{1/\varepsilon}}, \tag{16}$$

which is a useful form, because the RHS of (16) is composed solely of

[6] Subsequently we shall note parametric conditions under which this hypothesis is justified.

exogenous parameters of the economy — including what might be called 'initial conditions'.

We proceed now to simplify the equilibrium condition (8) by making use of (10) to express it as

$$n-\varepsilon=\frac{1-e^{-gnT}}{e^{gT}-1}\equiv\frac{1-e^{-gv}}{e^{gT}-1}. \tag{17}$$

With our approximations (17) reduces to

$$\frac{v}{T}-\varepsilon\simeq\frac{gv-(gv)^2/2}{gT+(gT)^2/2}\simeq\frac{v}{T}\left[\frac{1-gv/2}{1+gT/2}\right],$$

or

$$\frac{v}{T}\left[\frac{gT}{2}+\frac{gv}{2}\right]\simeq\varepsilon\left(1+\frac{gT}{2}\right)\simeq\varepsilon,$$

and, therefore that

$$gv^2\simeq 2\varepsilon T. \tag{18}$$

Finally, we analyse the determinants of the steady state rate of technological progress. Now we have noted that the ratio of market price to unit cost of production associated with the best-practice technique at each innovation date is the same. It follows from eq. (14) that the ratio of the cost of innovation to gross industrial revenue is also independent of the date of innovation. However, for the representative interval between innovations, $(0,T)$, we have $p_T=p_0e^{-gT}$ and $Q_T=Q_0e^{(m+g)T/\varepsilon}$. It follows then that for a steady state to be maintained

$$h\equiv(\mathrm{d}F_t/\mathrm{d}t)/F_t=-g+(m+g)/\varepsilon, \tag{19}$$

which, on using eq. (5) means that the steady state rate of technical progress, g, must be the solution of the equation

$$G(g)=-g+(m+g)/\varepsilon. \tag{20}$$

The steady state rate of technical progress is determined by $m$, $\varepsilon$ and the parameters underlying the innovation frontier $G(.)$. It is independent of $T$, and quite naturally, of initial conditions $F_0$ and $C_0$. (See fig. 1.)

We may now summarize the steady state equilibrium for the industry in question. As usual, let us concentrate our attention on the arbitrary innovation date $t=0$. The variables to be determined within the model are

$p_0$, $h$, $v$, $T$, $g$, $\dot{Q}_0$, $Q_0^i$ $[i=0,1,\ldots,(n-1)]$.[7] They are obtained from equations (1), (7), (10), (15), (18), (19) and (20). Aside from $\varepsilon$, $r$, $m$ and the parameters underlying the innovation frontier $G(.)$, $F_0$ and $C_0$ are exogenously given.

## 3. Characteristics of oligopoly equilibrium

An important variant of a classic question in industrial organization will be asked here. What is the relationship between the degree of concentration and the frequency of innovations? To see this, recall that by definition $v=nT$. Then eq. (18) reduces to the form

$$gn^2 \simeq 2\varepsilon/T. \tag{21}$$

It is tempting to use, say, the inverse of the number of active firms in the industry as an *index of concentration*. Eq. (21) says that *oligopoly equilibria are characterized by a negative association between this index of concentration and the frequency of innovations*. Thus, in a cross-section study of industries differing in their initial conditions, $F_0$ and $C_0$, one will observe such a negative association. But since $n$ and $T$ are both endogenous in the model such a relationship must not be given any *causal* interpretation: *industrial concentration and the frequency of innovations are simultaneously determined in the model.*

Typically though, the inverse of the number of firms in an industry has not been used as an index of concentration. Perhaps the most popular summary measure is the *concentration ratio*. For our model we may as well regard the concentration ratio as the ratio of sales of the most recent innovator to aggregate industry sales. From (6) and (10) note that

$$Q^0/Q \simeq gv/\varepsilon, \tag{22}$$

which, on using eq. (16), yields

$$Q^0/Q \simeq (2gT/\varepsilon)^{\frac{1}{2}}. \tag{23}$$

*Eq. (23) says that in a cross-section study of industries differing by their initial conditions $F_0$, $C_0$, there will be a negative association between the concentration ratio and the frequency of innovations. But again, this relationship must not be given any causal interpretation.*[8]

[7] $n$ is obtained from the identity $v \equiv nT$.

[8] These foregoing two results can be summarized by saying that so long as the degree of concentration is not too large, a higher level of industrial concentration is associated with less frequent innovations. But the *magnitude* of the innovations (i.e. cost reduction) at each date of innovation will be higher for the more concentrated industry.

We are particularly interested in the rate of technological progress $g$, and hence the growth in industrial output. For, at each innovation date industrial output increases by a factor $e^{gT/\varepsilon}$. The long-run average rate of growth of output is therefore $(m+g)/\varepsilon$. We turn now to eq. (20), which determines the steady state value of $g$. To obtain sharp answers suppose that the technological innovation frontier $G(g)$ is of the form

$$G(g)=\beta H(g), \tag{24}$$

where $H(g)$ is strictly increasing, strictly convex, with $H(0)=0$ and $H(\infty)=\infty$. Notice now from (24) and (20) that the condition $(m+g)/\varepsilon>r$ which, under our approximations, we have noted is necessary for the existence of a steady state equilibrium, is most certainly satisfied if $m$ is not too small. Thus suppose that $m$ is large enough to ensure that $(m+g)>r\varepsilon$. Let $g_c$ be the (unique) solution of eq. (20), and let $\bar{g}$ denote the economy-wide average rate of growth of output.[9] We may now dissect the determinants of industry growth in equilibrium and in turn identify parameters that determine the industry's performance relative to that of the rest of the economy. An industry which keeps par with the rest of the economy is one for which $(m+g_c)/\varepsilon=\bar{g}$, for in this case the long-run average growth in industry output is the same as the rest of the economy. In the obvious manner one may define a 'declining' or 'growing' industry — relative to the rest of the economy. Notice now that $\partial g_c/\partial m>0$, $\partial g_c/\partial\varepsilon<0$ and $\partial g_c/\partial\beta>0$. Therefore, *industries facing a higher exogenous growth rate in the demand for their product experience a faster long-run rate of technical progress and in turn a higher long-run rate of growth in output.*[10] More interestingly, perhaps, *industries facing a more elastic demand for their products experience a faster long-run rate of technical progress and also a higher long-run rate of growth in output* [since $\partial((m+g_c)/\varepsilon)/\partial\varepsilon<0$]. The model implies that if two industries are to remain on par with each other in terms of long-run growth performance, then cet. par. the exogenous growth in the demand for the product of the industry facing the more inelastic demand must be greater. Finally, *industries that are better endowed for technological progress (lower value of $\beta$) enjoy a higher long-run rate of growth in output.*

We may in fact solve the equilibrium conditions explicitly. Thus, on using eq. (18) in eq. (16) we obtain the equilibrium period between innovations, $T_c$, as

$$T_c\simeq\left(\frac{F_0C_0^{(1-\varepsilon)/\varepsilon}}{2\varepsilon((m+g_c)/\varepsilon-r)\sigma^{1/\varepsilon}}\right)^{\frac{1}{2}}, \tag{25}$$

[9] In what follows the subscript $c$ below a variable denotes the value of the variable under competitive equilibrium. The reader may confirm that if $F_0$ is sufficiently small and $(m+g_c)>r\varepsilon$, then a steady state equilibrium exists and is in fact the state we are analysing here.

[10] See Schmookler (1962).

which, on using in (18) yields the equilibrium life of a firm, $v_c$, as

$$v_c^2 \simeq 2\varepsilon T_c/g_c. \tag{26}$$

Using eq. (26) in (18) we may obtain the equilibrium number of firms, $n_c$, as

$$n_c^2 \simeq 2\varepsilon/T_c g_c. \tag{27}$$

Finally, an appeal to eq. (22) yields the concentration ratio as

$$Q_c^0/Q_c \simeq (2g_c T_c/\varepsilon)^{\frac{1}{2}}. \tag{28}$$

*It follows immediately that an increase in (initial) innovation costs, $F_0$, reduces the frequency with which innovations are undertaken, reduces the equilibrium number of firms and at the same time increases the life span of the representative firm. It also results in an increase in the concentration ratio.* These conclusions are intuitively reasonable; for innovation costs are a cost of entry, and these results make clear that to the extent innovation costs can be affected by government policy (e.g. by development subsidies), industrial structure can be affected in ways that one might expect.

Now quite clearly government policy can directly influence $C_0$, e.g. by a constant rate of ad-valorum tax or subsidy on each unit produced. *Eqs. (25)–(28) make the somewhat surprising points that if $\varepsilon < 1$ a reduction in unit operating cost, $C_0$, increases the frequency of innovations, reduces the life of the representative firm, increases the number of firms in the industry and decreases the concentration ratio. Just the reverse is the case if demand is inelastic; i.e. $\varepsilon > 1$, as eqs. (25)–(28) make clear.* It is curious that the effect of a change in the unit cost of production on the characteristics of an oligopoly equilibrium depends so critically on the elasticity of demand.

The effect of an increase in the exogenous growth rate in demand, $m$, is somewhat predictable. *Industries experiencing a greater growth in the market for their products undertake innovations more frequently and the lifespan of the representative firm in such industries is shorter.* So too are the effects of the interest rate on industrial structure in line with what one might expect. *An increase in the interest rate, $r$, reduces the frequency of innovations, increases the lifespan of the representative firm, reduces the equilibrium number of firms in the industry and causes an increase in the concentration ratio.* Finally, *an increase in the size of the market, $\sigma$, increases the frequency of innovations, reduces the lifespan of the representative firm, increases the equilibrium number of firms in the industry, and causes a decline in the concentration ratio.*

## 4. The socially managed industry

We turn now to an analysis of the socially managed industry. To keep matters tractable we suppose that the planner chooses among steady states. It is clear that the planner will use only the existing best-practice technique; i.e. only the latest vintage will be in operation at any time. We suppose as well that the government can finance its expenditures from general taxation. Hence social welfare maximization dictates the marginal-cost pricing rule. Thus $p_t = C_t$ for all $t$, which, on using (1), implies that

$$Q_t = [\sigma/C_t]^{1/\varepsilon} e^{(m/\varepsilon)t}. \tag{29}$$

We may now note that the hypothesis that the socially managed industry is also in a steady state implies that the rate of technical progress, $g_s$, for this industry is, like the competitive one, the solution of eq. (20). Thus $g_c = g_s$. This follows from the fact that along a steady state the ratio of innovation costs to gross revenue must be independent of the date of innovation. It follows that the long-run average growth rate in output in the socially managed industry is the same as that in the competitive one. Let $T$ denote the interval between innovations. Then we may note from (29) that in the intervals between innovations output grows at the rate $m/\varepsilon$ and that at each innovation date output increases discontinuously by the factor $e^{(g_s/\varepsilon)T}$. We are however interested in the optimum steady state; that is, the optimum value of $T$. Let us suppose that the competitive interest rate, $r$, is regarded as appropriate for discounting social benefits.

Now, the social utility function associated with the (compensated) demand curve (1) is

$$u(Q_t) = [\sigma/(1-\varepsilon)] Q_t^{(1-\varepsilon)} e^{mt}. \tag{30}$$

Thus suppose without loss of generality that $t = 0$ is an innovation date. Then if $F_0$ is the innovation cost and if $C_0$ is unit operating cost associated with the best-practice technique at $t = 0$, then during the interval $(0, T)$ the planner will set $p_t = C_0$, and so the flow of net social benefits, excluding innovation costs, during $(0, T)$ will be

$$u(Q_t) - C_0 Q_t = [\varepsilon/(1-\varepsilon)] \sigma^{1/\varepsilon} C_0^{-(1-\varepsilon)/\varepsilon} e^{(m/\varepsilon)t}, \tag{31}$$

and, more generally, the flow of net social benefits, excluding innovation costs, during the interval $(iT, (i+1)T)$, will be

$$u(Q_t) - C_0 e^{ig_s T} Q_t = [\varepsilon/(1-\varepsilon)] \sigma^{1/\varepsilon} C_0^{-(1-\varepsilon)/\varepsilon} e^{ig_s(1-\varepsilon)T/\varepsilon} e^{(m/\varepsilon)t}, \quad i = 0, 1, 2, \ldots. \tag{32}$$

It follows that the present discounted value of net social benefits, excluding innovation costs, is

$$[\varepsilon/(1-\varepsilon)]\sigma^{1/\varepsilon}C_0^{-(1-\varepsilon)/\varepsilon}\left[\frac{1-e^{-(r-m/\varepsilon)T_s}}{(r-m/\varepsilon)}\right]\sum_{i=0}^{\infty} e^{-i(r+g_s-(m+g_s)/\varepsilon)T}, \tag{33}$$

and we assume that $r+g_s>(m+g_s)/\varepsilon$.[11]

We turn now to the present value of innovation costs. Along a steady state innovation costs increase at the percentage rate $h=-g_s+(m+g_s)/\varepsilon$. Therefore, the present value of the flow of innovation expenditure is

$$F_0\sum_{i=0}^{\infty} e^{-i[r+g_s-(m+g_s)/\varepsilon]T}. \tag{34}$$

Write

$$r-m/\varepsilon\equiv a, \qquad r+g_s-(m+g_s)/\varepsilon\equiv b,$$
$$[\varepsilon/(1-\varepsilon)]\sigma^{1/\varepsilon}C_0^{-(1-\varepsilon)/\varepsilon}\equiv A. \tag{35}$$

Combining (33) and (34) we may now express the present value of the flow of net social benefits (inclusive of innovation costs) as

$$\left[\frac{A}{a}(1-e^{-aT})-F_0\right](1-e^{-bT})^{-1}. \tag{36}$$

The social planner's problem consists in choosing $T$ with a view to maximizing (36). The first-order condition associated with this problem then is

$$Ae^{-aT}-be^{-bT}\left[\frac{A}{a}(1-e^{-aT})-F_0\right][1-e^{-bT}]^{-1}=0. \tag{37}$$

To simplify suppose again that $F_0$ is small, so that the optimal value of $T$, call it $T_s$, is also small. In this case eq. (37) can be approximated as

$$A(a-b)T^2+bF_0T-F_0\simeq 0, \tag{38}$$

and so

$$T_s\simeq[F_0/A(a-b)]^{\frac{1}{2}}=\left[\frac{F_0C_0^{(1-\varepsilon)/\varepsilon}}{g_s\sigma^{1/\varepsilon}}\right]^{\frac{1}{2}}. \tag{39}$$

[11]Thus, when we come to comparing the competitive industry with the socially managed one we must suppose both $m+g_c>r\varepsilon$ and $(r+g_s)\varepsilon>m+g_s$; i.e. that $m+g_c>r\varepsilon>m+g_s-\varepsilon g_s$ (with $g_c=g_s$).

*We conclude immediately that the optimal frequency of innovations is greater in industries facing a larger growth in demand, (m); is a decreasing function of the initial cost of innovation, is an increasing function of the size of the market, $\sigma$, and is a decreasing or increasing function of the initial operating cost depending on whether or not the market demand curve is elastic. Most remarkable, perhaps, eq. (39) tells us that to a first approximation the optimum frequency of innovations is independent of the social rate of discount.*

The question arises how the competitive frequency of innovation compares with the socially optimal frequency. *To see this one must compare eqs. (25) and (39) to establish that* $T_c \gtreqless T_s$ *as* $2\varepsilon((m+g_c)/\varepsilon)-r) \gtreqless g_s$, *where* $g_c(=g_s)$ *is the (unique) solution of eq. (20).*

*In particular, this result tells us that competition encourages innovations to be undertaken too frequently if the exogenous rate of growth in demand, m, is large.*[12] Thus, a competitive industry facing a large growth in demand will be characterized by 'small' innovations occurring too frequently.

## 5. Pure monopoly

The case of the pure monopolist, protected fully by entry barriers, parallels the analysis presented in the previous section. But we must now restrict ourselves to the case $0<\varepsilon<1$. We may therefore merely present the results without supplying details. In what follows we suppose that the monopolist faces a perfect capital market, where the interest rate is $r$. The monopolist, we assume, wishes to maximize the present value of net profits, and that his choice is restricted to one among steady states. This latter assumption implies as before that the rate of technical progress, call it $g_m$, that the monopolist faces is the solution of eq. (20). Thus $g_c = g_s = g_m$. Without loss of generality let us suppose that the monopolist innovates at $t=0$. Then it follows from (1) that the monopolist's output, $Q_0$, at this date is

$$Q_0 = [\sigma(1-\varepsilon)/C_0]^{1/\varepsilon}. \tag{40}$$

Write

$$\begin{aligned} &r - m/\varepsilon \equiv a, \qquad r + g_m - (m+g_m)/\varepsilon \equiv b, \\ &(1-\varepsilon)^{1/\varepsilon}[\varepsilon/(1-\varepsilon)]\sigma^{1/\varepsilon} C_0^{-(1-\varepsilon)/\varepsilon} \equiv B. \end{aligned} \tag{41}$$

Suppose $T$ is the interval between innovations. Then an argument identical to the one leading to (36) establishes the the monopolist will wish to choose $T$ with a view to maximizing the expression

$$\left[\frac{B}{a}(1-e^{-aT}) - F_0\right][1-e^{-bT}]^{-1}. \tag{42}$$

[12]In particular, if $m > \varepsilon r$.

As before, we approximate and suppose that $F_0$ is 'small'. Let $T_m$ denote the monopolist's profit maximizing choice of the interval between innovations. Then, an argument parallel to the one leading to eq. (39) establishes that

$$T_m \simeq \left[\frac{F_0 C_0^{(1-\varepsilon)/\varepsilon}}{g_m\{\sigma(1-\varepsilon)\}^{1/\varepsilon}}\right]^{\frac{1}{2}}. \tag{43}$$

*If we now compare eqs. (39) and (43) we conclude that a monopolist, protected fully by entry barriers, undertakes innovations less frequently than is socially optimal. However, the monopolist's innovations are always bigger than the socially optimal ones* (this follows from the fact that $g_s = g_m$).

## 6. Commentary and extension

Our principal aim in this paper has been to construct a mode of analysis for locating the determinants of the long-run growth in output of an industry, and thereby to seek an explanation of the performance of an industry relative to that of others. Towards this end we have supposed that the spurt to an industry's growth has as its source an exogenous increase in the demand for its product, and a steady supply of basic ideas and inventions that can be exploited — at a cost, of course — for the purposes of process innovation. Our interest in the long-run performance of an industry *suggests* that the analysis be restricted to that of steady states. Equally important, the theoretical construct of a steady state makes the model tractable. Since innovations involve fixed costs we know in advance that they will not occur continuously — nor is it desirable that they do so. So the important question arises as to the frequency with which innovations occur under alternative industrial organizations and also the magnitude of the innovations when they do occur. It is important to recognize that industrial organizations may differ with respect to the frequency and magnitude of innovations undertaken within them even while they are identical with respect to the long-run growth in output that they sustain. The model we have analyzed in this paper demonstrates this sharply, for we showed that the model implies that the long-run growth in output along a steady state is independent of industrial organization.

Much of the analysis of this paper has been directed at a competitive industry (i.e. one characterized by free entry and non-cooperative behaviour). It is the opportunity of earning profits which ensures that innovations occur at steady intervals, but competition drives the profits obtained by the innovating firms to zero. Since innovations involve fixed costs, competition among firms results in an oligopolistic structure where, at any instant, only a finite number of firms are actually engaged in production. Entry by a firm that has innovated results in a fall in product price — and hence an increase in industry output — which in turn results in the economic obsolescence of the most ageing machines. For our model, this means the exit of the oldest firm in operation. A major weakness of the competitive model, of sections 2 and 3, as we see it, is the hypothesis that a given firm is entitled to innovate at most once. This can be justified by postulating the presence of strong anti-trust legislation which deters existing firms from undertaking entry-deterring innovations. Removing this assumption greatly complicates the analysis.

The advantages afforded by our insistence that the industry under study is in a steady state are obvious. Equally so, perhaps, are the disadvantages. We do not wish to make the vulgar complaint that steady-states are only a myth, but rather that the assumption really bolts down the rate of technical change for the industry and therefore the long-run growth in output — as eq. (20) makes clear. Obviously, one would like to make the rate of technical change, $g$, a *choice* variable, and the steady state hypothesis prevents us from doing that, *unless* the initial conditions, $F_0$ and $C_0$, have, by fluke, exactly the 'appropriate' values. To see this let us return to the innovation frontier given by eq. (2). Assume that the economy is in a steady state. Let us suppose that at $t$ a firm contemplates on entering the industry at $t+\theta$, and suppose we allow the firm to choose a pair $(C_{t+\theta}, F_{t+\theta})$ from the feasible set $\phi(C_t, F_t)$. Then clearly it would wish to choose with a view to minimizing the discounted value of its total costs. Now let $Z_{t+\theta}(r, g, m, T, n)$ denote the sum, computed at $t+\theta$, of the discounted values of the flow of *output* of the innovating firm during its lifetime, $(t+\theta, t+T+\theta)$. Then the firm wishing to innovate at $t+\theta$ must choose $C_{t+\theta}$ so as to minimize $C_{t+\theta}Z_{t+\theta}+F_{t+\theta}$, which on using eq. (2) becomes

$$C_{t+\theta}Z_{t+\theta}+F_t[1+G((C_t-C_{t+\theta})/\theta C_{t+\theta})\theta)]. \tag{44}$$

The first-order condition associated with (44) is

$$Z_{t+\theta}-C_tF_tG'(\,.\,)/\theta C_{t+\theta}^2=0,$$

which, on moving to continuous time by letting $\theta \to 0$ reduces to the condition

$$Z_tC_t/F_t=G'(g). \tag{45}$$

Now return to the case of a firm innovating at $t=0$. Routine calculations show that along a steady state

$$Z_0=Q_0^0\left[\frac{1-e^{-(r-m/\varepsilon)T}}{(r-m/\varepsilon)}\right]\sum_{i=0}^{n-1} e^{-i(r-(m+g)/\varepsilon)T}, \tag{46}$$

where $Q_0^0$ is the output of the innovating firm at $t=0$. But for small $F_0$, and therefore small $T$, (46) reduces to

$$Z_0 \simeq Q_0^0 v,$$

which, on using in eq. (45) at $t=0$, means that

$$C_0 Q_0^0 v / F_0 \simeq G'(g). \tag{47}$$

But along a steady state, eq. (20) must also hold, and $v$ and $Q_0^0$ must satisfy eqs. (26) and (28). The system is therefore overdeterminate, unless $F_0$ and $C_0$, by fluke, have precisely those values which ensure that the value of g which solves eq. (47) also satisfies eq. (20). However, it should be noted that if the forces of competition, in which the rate of technical progress is subject to choice, drives the industry in the long run to a steady state — and we can present no theory which says it does — then it must be the case that the cost structure (i.e. $F_t$ and $C_t$) must tend in the long run to that locus which makes eqs. (20) and (47) consistent.

We shall not attempt to summarize each of the results in this paper, but rather shall emphasize the fact that the model analyzed here suggests that if the growth in demand for the product of an industry is high the forces of competition result in too frequent a set of innovations, none of which on its own is sufficiently large; and that there are reasons for believing that a monopolist, protected fully by entry barriers, undertakes innovations less frequently than is socially desirable, but that when it does, it undertakes unduly large innovations. It is not clear what intuition one should have had about the nature of these biases.

## Appendix

Here we want to show that for 'small' values of $T$ and $v$, the zero-profit condition (14) reduces to eq. (15) in the text. Now, it will be noted that eq. (14) can be expressed as

$$\frac{p_0 Q_0[1-e^{-(r-m/\varepsilon)T}]}{(r-m/\varepsilon)}\left[\frac{1-e^{-n[r+g-(m+g)/\varepsilon]T}}{1-e^{-[r+g-(m+g)/\varepsilon]T}}\right.$$
$$\left.+\frac{C_0^2(1-e^{-n[r-g-(m+g)/\varepsilon]T})}{p_0^2(1-e^{-[r-g-(m+g)/\varepsilon)T]}}-\frac{2C_0(1-e^{-n[r-(m+g)/\varepsilon]T})}{p_0(1-e^{-[r-(m+g)/\varepsilon]T})}\right]=F_0. \tag{A.1}$$

Now use eqs. (1) and (10) in the text in (A.1) to obtain

$$\frac{(1-\mathrm{e}^{-K_1T})}{\varepsilon K_1}\left[\frac{1-\mathrm{e}^{-K_2v}}{1-\mathrm{e}^{-K_2T}}+\mathrm{e}^{-2gv}\frac{(1-\mathrm{e}^{-K_3v})}{(1-\mathrm{e}^{-K_3T})}-2\mathrm{e}^{-gv}\frac{(1-\mathrm{e}^{-K_4v})}{(1-\mathrm{e}^{-K_4T})}\right]$$

$$=F_0/p_0Q_0=\frac{F_0C_0^{(1-\varepsilon)/\varepsilon}\mathrm{e}^{gv(1-\varepsilon)/\varepsilon}}{(1-\varepsilon)^{1/\varepsilon}}, \tag{A.2}$$

where

$$K_1\equiv r-m/\varepsilon,$$

$$K_2\equiv r+g-(m+g)/\varepsilon,$$

$$K_3=r-g-(m+g)/\varepsilon,$$

$$K_4=r-(m+g)/\varepsilon.$$

Now, if $K$ is a constant, then for small $v$ and $T$ (with $v/T$ large) we have

$$\begin{aligned}\frac{1-\mathrm{e}^{-Kv}}{1-\mathrm{e}^{-KT}}&\simeq\frac{Kv-(Kv)^2/2}{KT-(TK)^2/2}=\frac{v}{T}\left(\frac{1-vK/2}{1-TK/2}\right)\\&\simeq\frac{v}{T}(1-vK/2)(1+TK/2)\simeq\frac{v}{T}(1-vK/2).\end{aligned} \tag{A.3}$$

If we now use (A.3) in the terms within the square brackets of (A.2) we have

$$\begin{aligned}&\frac{1-\mathrm{e}^{-K_2v}}{1-\mathrm{e}^{-K_2T}}+\mathrm{e}^{-2gv}\frac{(1-\mathrm{e}^{-K_3v})}{(1-\mathrm{e}^{-K_3T})}-2\mathrm{e}^{-gv}\frac{(1-\mathrm{e}^{-K_4v})}{(1-\mathrm{e}^{-K_4T})}\\
&\simeq\frac{v}{T}(1+\mathrm{e}^{-2gv}-2\mathrm{e}^{-gv})-\frac{v^2}{2T}(K_2+K_3\mathrm{e}^{-2gv}-2K_4\mathrm{e}^{-gv})\\
&\simeq\frac{v}{T}(1-\mathrm{e}^{-gv})^2-\frac{v^2}{2T}(K_2+K_3\mathrm{e}^{-2gv}-2K_4\mathrm{e}^{-gv})\\
&\simeq\frac{v}{T}(gv-g^2v^2/2)^2-\frac{v^2}{2T}(K_3(-2gv+2g^2v^2)-2K_4(-gv+g^2v^2/2))\\
&\qquad\qquad(\text{since}\quad K_2+K_3-2K_4=0)\\
&\simeq g^2\frac{v^3}{T}(1-gv/2)^2-\frac{v^2}{2T}(-2gvK_3+2gvK_4+(2K_3-K_4)g^2v^2)\\
&\simeq\frac{g^2v^4}{2T}\left(\frac{(m+g)}{\varepsilon}-r\right).\end{aligned} \tag{A.4}$$

On using (A.4) and (A.2) one obtains eq. (15) in the text.

## References

Dasgupta, P. and J. Stiglitz, 1977, Innovations and market structure, Mimeo. (London School of Economics, London).

Dasgupta, P. and J. Stiglitz, 1980a, Industrial structure and the nature of innovative activity, Economic Journal 90, 266–293.

Dasgupta, P. and J. Stiglitz, 1980b, Uncertainty, industrial structure and the speed of R&D, Bell Journal of Economics 11, Spring, no. 1, 1–28.

Dasgupta, P., R. Gilbert and J. Stiglitz, 1980a, Invention and innovation under alternative market structures: The case of natural resources, Mimeo. (Princeton University, Princeton, NJ).

Dasgupta, P., R. Gilbert and J. Stiglitz, 1980b, Energy resources and research and development, in: H. Siebert, ed., Erschöpfbare Ressourcen (Duncker & Humblot, Berlin).

Schmookler, J., 1962, Economic sources of innovative activity, Journal of Economic History 22.